Parvovirus-derived endogenous viral elements in two South American rodent genomes.

Gloria Arriagada[1*] and Robert J. Gifford[2*]

[1]Universidad Andres Bello, Departamento de Ciencias Biologicas, Facultad de Ciencias Biologicas, Quillota 980, Viña del Mar, Chile

[2]MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

[*]Corresponding authors

Abstract

We describe endogenous viral elements (EVEs) derived from parvoviruses (family *Parvoviridae*) in the long-tailed chinchilla (*Chinchilla lanigera*) and degu (*Octodon degus*) genomes. The novel EVEs include Dependovirus-related elements, and representatives of a clearly distinct parvovirus lineage that also has endogenous representatives in marsupial genomes. In the degu, one dependovirus-derived EVEs was found to encode an intact reading frame, and was differentially expressed *in vivo*, with increased expression in the liver.

Main text

Parvoviruses are small, non-enveloped viruses containing a single-stranded DNA (ssDNA) genome ~5 kilobases (kb) in length. In recent years, several reports have described endogenous viral elements (EVEs) derived from parvoviruses in animal genomes (1-4). We performed an *in silico* screen of recently published low-coverage genome sequence assemblies using the *database-integrated genome screening* (DIGS) tool (5). Screening identified novel parvovirus-related EVEs in two caviomorph rodents - the long-tailed chinchilla (*Chinchilla lanigera*), and the degu (*Octodon degus*). A total of twelve novel EVEs were identified in these two species (**Table 1**).

Chinchillas are medium-sized, crepuscular rodents that live at high altitude in the Andes mountains. Degus are small rodents endemic to the Chilean Matorral ecoregion. Although both species are indigenous to South America, they are relatively distantly related, having diverged ~37 million years ago (MYA) (6).

Parvovirus genomes generally comprise two major gene cassettes that separately encode nonstructural (NS) and structural (VP) proteins. Genetic structures of previously described parvovirus-EVEs have included complete viral genomes, intact individual genes, isolated genome fragments, and rearranged complete genomes (2, 3). A similar range of genetic structures were observed here, with five of the novel EVEs representing complete or nearly complete viral genomes spanning both major gene cassettes, while the remainder represented single gene cassettes, or fragments of genes (**Figure 1**).

Phylogenetic analysis of the newly identified EVEs revealed that half of them grouped robustly within the diversity of avian and mammalian Dependoviruses, whereas the others clustered in a single, well-defined clade comprised exclusively of EVEs obtained from the genomes of South American and Australian mammals (**Figure 2**). In phylogenies, this lineage of EVEs formed a sister clade with exogenous viruses in the Parvovirus and 'Bufavirus' (7) genera.

None of the novel EVEs were orthologous in the species examined, thus we could not infer minimum dates of integration based on orthology. However, we did identify a pair of endogenous Dependovirus elements (*C.lanigera* 5a and 5b (see **Table 1**)) that had apparently been duplicated after integrating into the chinchilla genome. This pair of nearly identical elements shared at least two nonsense mutations (**Figure S1**), indicating that the coding sequence had degenerated prior to its duplication, and can thus be assumed to have been evolving neutrally in the subsequent period. These elements are at least as old as the duplication event that generated them, which we estimated to have arisen in a duplication event that occurred between 2.8 and 5.6 MYA (**Figure S1**).

Although most of the novel EVEs contained frameshifts and/or stop codons, insertions encoding apparently intact NS1 proteins were identified in both the chinchilla and

degu genomes. To exclude the possibility that it was somehow derived from contaminating viral DNA (8, 9), one intact element (*O.degus*-4) was independently amplified from genomic DNA by polymerase chain reaction (PCR). Tissue was obtained from a fresh male headless *O.degus* cadaver (kindly donated by Dr. Adrian Palacios from Universidad de Valparaiso, Chile). Genomic DNA was extracted from liver tissue, and PCR using primers targeting the 5' and 3' flanking regions of *O.degus*-4 (**Figure 3a**), confirmed the presence of this EVE locus in a second, outbred degu individual, demonstrating that *O.degus*-4 is a genuinely endogenous sequence.

RT-PCR was used to investigate expression of *O.degus*-4 in distinct degu tissues. RNA was extracted from pancreas, liver, testicle, kidney, suprarenal, spleen and lung tissues (**Figure 3b**). This analysis revealed that the *O.degus*-4 replicase is differentially expressed *in vivo*, with markedly elevated expression of mRNA in the liver, and little or no expression in other tissues.

Scientifically, EVEs can be approached from two overlapping but distinct perspectives. Firstly, they can be viewed a kind of genomic 'fossil record' from which the long-term, coevolutionary relationships of viruses and hosts can be inferred. In this respect, the presence of a specific, monophyletic lineage of EVEs in both South American and Australian marsupial genomes -and the apparent absence of this lineage from the genomes of Old World rodents - suggests the existence of an ancient parvovirus lineage that evolved in the indigenous mammal populations of biogeographically isolated Southern hemisphere continents (marsupials, xenarthrans), and was acquired by caviomorph rodents subsequent to their colonization of the South American continent (estimated to have occurred ~40 MYA (10)).

The second way in which EVEs can be viewed is as host genes. While most EVE sequences are highly degenerated, it is clear that at least a proportion of these elements have been co-opted or 'exapted' (i.e. adapted for a function distinct from that for which they originally evolved (11)) to perform physiological functions in their host species (12-15). Dependovirus-derived EVEs encoding intact replicase proteins have previously been

identified in mammalian genomes, including the African elephant (*Loxodonta africana*) and Hamadryas baboon (*Papio hamadryas*). However, this is the first study to show that a parvovirus-derived EVE is expressed *in vivo* in a mammal. While no physiological function has yet been demonstrated for the numerous Parvovirus-related EVEs in mammalian genomes, the identification of an intact element with differential expression across tissues provides further indication that such functions exist. Since degus are experimental organisms that are used currently to research mammalian pathologies and behaviours (16, 17), the identification of an intact, expressed parvovirus-derived EVE in this species suggests a possible path forward for research in this area.

Figure legends

Figure 1.

(a) Genetic structures of parvovirus-derived EVEs in *C.lanigera* and *O.degus* genomes. ORFs were inferred by manual comparison of putative peptide sequences to those of closely related exogenous parvoviruses.

(b) Alignment of a pair of duplicated EVEs in the *Chinchilla lanigera* genome against an adeno-associated virus type 2 (AAV2) reference sequence (GenBank accession #: AF043303.1). Blocks of six or more amino acids that are conserved between all three sequences are highlighted in grey. Nonsense mutations shared between the two duplicated EVEs are highlighted in black. The two EVEs differed at 8 of 323 nucleotide positions (i.e. an average of 4 nucleotide substitutions has occurred in each insertion), indicating they arose in a duplication event that occurred between 2.8 and 5.6 MYA (assuming a range of mammalian neutral substitution rates from 2.2 to 4.5 x 10-9 per site per year (18)).

Figure 2.

Maximum likelihood phylogenies showing the relationships of **(a)** parvovirus-related, and **(b)** dependovirus-related EVEs in the *C.lanigera* and *O.degus* genomes to exogenous parvoviruses and previously described EVEs. Phylogenies are based on alignments of NS1 proteins and putative NS1 proteins encoded by EVE pseudogenes, and were constructed using PHYML (19) and the JTT+ protein substitution model as selected by ProtTest (20). EVEs identified in this study are underlined. The scale bar indicates evolutionary distance in substitutions per amino acid site. Asterisks indicate nodes with maximum likelihood bootstrap support above 75% for 100 bootstrap replicates. Abbreviations: EVE=endogenous viral element; AAV=adeno-associated virus; MV= minute virus; PV=Parvovirus. NS=non-structural protein gene cassette; VP=viral capsid protein gene cassette.

Figure 3.

(a) Schematic representation of endogenous viral element (EVE) *O.degus*-4. The replicase ORF is shown as a white box and the 5' and 3' genomic sequences as black lines. The relative positions of primers used in this study are shown. Primers F1 and F2, positioned at the extreme ends of the EVE were used to amplify genomic DNA and sequence PCR products. Primers G1 and G2, positioned in the genomic flanking sequence, were used to control for genomic contamination of the RNA preparations. The internal primers I1 and I2 were used to detect *O.degus*-4 replicase mRNA.

(b) Tissue-specific expression of a dependovirus-related EVE mRNA in the degu. Total RNA was extracted from *O.degus* pancreas, lung, testicle, suprarenal gland, spleen, liver and kidney. After cDNA generation, expression of the *O.degus*-4 EVE was detected using primers designed to amplify the region 519-689 of the predicted mRNA (top panel). OdGAPDH mRNA (XM_004643553.1) was used as a positive control of mRNA presence (middle panel). To discard genomic DNA contamination in the cDNA preparations, primers aligning 35 nucleotides (nt) upstream and 85 nt downstream the predicted element in the genomic sequence NW_004524640.1 were used (bottom panel). The different tissues analyzed are indicated above the figure. Genomic DNA was used as a positive amplification control.

Acknowledgements

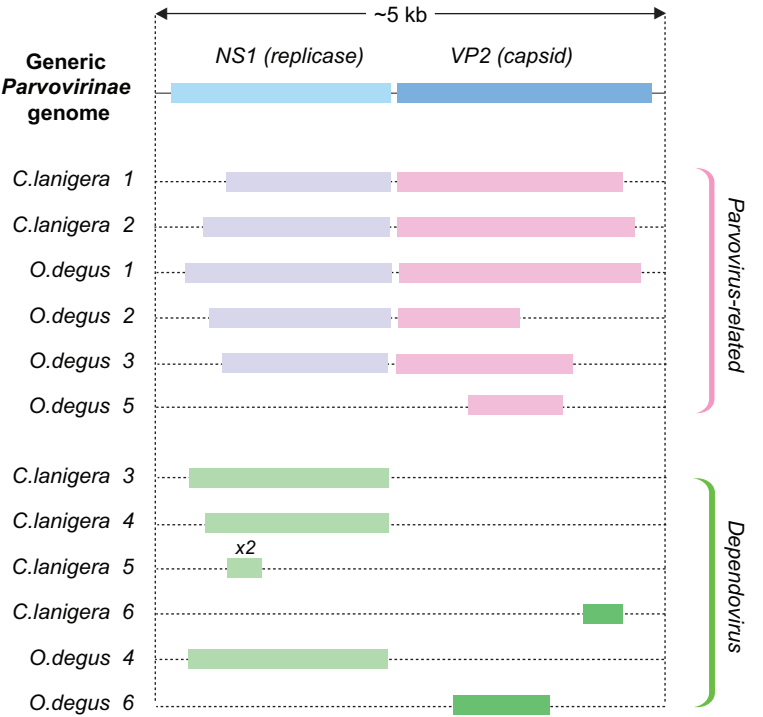
This work was supported by FONDECYT 1130852 (G.A.), and UK Medical Research Council grant number MC_UU_12014 (R.G.).

References

1. **Belyi VA, Levine AJ, Skalka AM.** 2010. Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are more than 40 to 50 million years old. J Virol **84:**12458-12462.
2. **Kapoor A, Simmonds P, Lipkin WI.** 2010. Discovery and characterization of mammalian endogenous parvoviruses. J Virol **84:**12628-12635.

3. **Katzourakis A, Gifford RJ.** 2010. Endogenous viral elements in animal genomes. PLoS Genet **6:**e1001191.
4. **Liu H, Fu Y, Xie J, Cheng J, Ghabrial SA, Li G, Peng Y, Yi X, Jiang D.** 2011. Widespread endogenization of densoviruses and parvoviruses in animal and human genomes. J Virol **85:**9863-9876.
5. **Gifford RJ.** 2014. DIGS v1.0, Available from http://paleovirology.org.uk/.
6. **Opazo JC.** 2005. A molecular timescale for caviomorph rodents (Mammalia, Hystricognathi). Molecular phylogenetics and evolution **37:**932-937.
7. **Phan TG, Vo NP, Bonkoungou IJ, Kapoor A, Barro N, O'Ryan M, Kapusinszky B, Wang C, Delwart E.** 2012. Acute diarrhea in West African children: diverse enteric viruses and a novel parvovirus genus. J Virol **86:**11024-11030.
8. **Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, Aronsohn A, Hackett J, Jr., Delwart EL, Chiu CY.** 2013. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. J Virol **87:**11966-11977.
9. **Smuts H, Kew M, Khan A, Korsman S.** 2014. Novel Hybrid Parvovirus-Like Virus, NIH-CQV/PHV, Contaminants in Silica Column-Based Nucleic Acid Extraction Kits. J Virol **88:**1398.
10. **Antoine PO, Marivaux L, Croft DA, Billet G, Ganerod M, Jaramillo C, Martin T, Orliac MJ, Tejada J, Altamirano AJ, Duranthon F, Fanjat G, Rousse S, Gismondi RS.** 2012. Middle Eocene rodents from Peruvian Amazonia reveal the pattern and timing of caviomorph origins and biogeography. Proceedings. Biological sciences / The Royal Society **279:**1319-1326.
11. **Gould SJ, Vrba ES.** 1982. Exaptation-A Missing Term in the Science of Form. Paleobiology **8:**4-15.
12. **Arnaud F, Caporale M, Varela M, Biek R, Chessa B, Alberti A, Golder M, Mura M, Zhang YP, Yu L, Pereira F, Demartini JC, Leymaster K, Spencer TE, Palmarini M.** 2007. A paradigm for virus-host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses. PLoS Pathog **3:**e170.
13. **Bezier A, Annaheim M, Herbiniere J, Wetterwald C, Gyapay G, Bernard-Samain S, Wincker P, Roditi I, Heller M, Belghazi M, Pfister-Wilhem R, Periquet G, Dupuy C, Huguet E, Volkoff AN, Lanzrein B, Drezen JM.** 2009. Polydnaviruses of braconid wasps derive from an ancestral nudivirus. Science **323:**926-930.
14. **Rowe HM, Trono D.** 2011. Dynamic control of endogenous retroviruses during development. Virology **411:**273-287.
15. **Dewannieux M, Heidmann T.** 2013. Endogenous retroviruses: acquisition, amplification and taming of genome invaders. Curr Opin Virol **3:**646-656.
16. **Okanoya K, Tokimoto N, Kumazawa N, Hihara S, Iriki A.** 2008. Tool-use training in a species of rodent: the emergence of an optimal motor strategy and functional understanding. PloS one **3:**e1860.
17. **Ardiles AO, Tapia-Rojas CC, Mandal M, Alexandre F, Kirkwood A, Inestrosa NC, Palacios AG.** 2012. Postsynaptic dysfunction is associated with spatial and object recognition memory loss in a natural model of Alzheimer's disease. Proceedings of the National Academy of Sciences of the United States of America **109:**13835-13840.
18. **Kumar S, Subramanian S.** 2002. Mutation rates in mammalian genomes. Proceedings of the National Academy of Sciences of the United States of America **99:**803-808.
19. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology **59:**307-321.
20. **Abascal F, Zardoya R, Posada D.** 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics (Oxford, England) **21:**2104-2105.

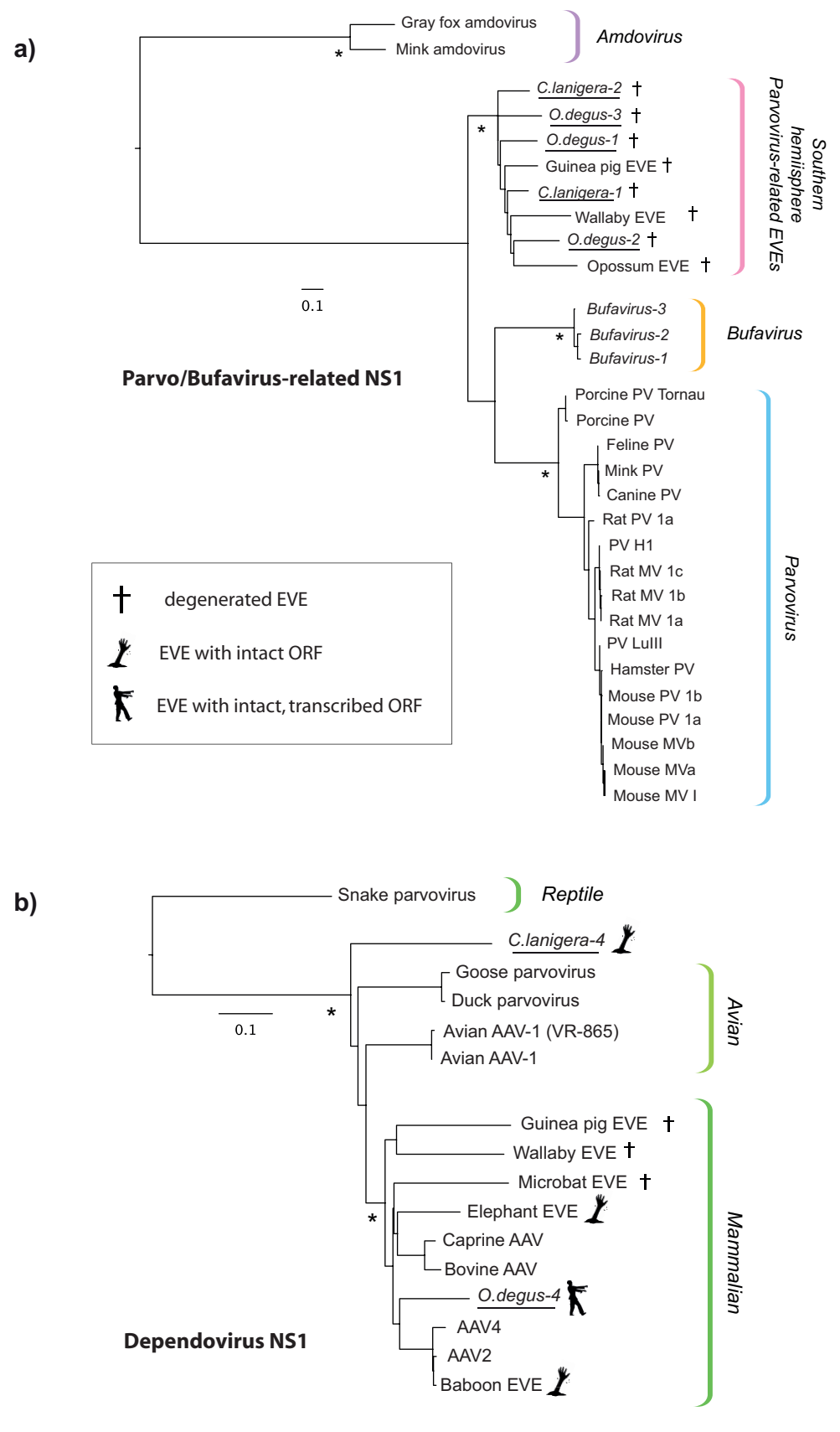
a)



b)

```
Chinchilla 5a   NKNFPFS?PLDKMIVWWEEGKMTAKVVESLTATLGSSKVRVDEKGKSSDELDQSPA
Chinchilla 5b   NKNFPFS?PLDKMIVWWEEGKITAKVVESLTASVGSSKVRVDEKGKSSDELDQSPA
AAV2            NENFPFNDCVDKMVIWWEEGKMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPV

Chinchilla 5a   IVIRNTDMCTLVDGN*TALEHQQCLQDWMSTFQLAYSLELSFGKITKQEVKD
Chinchilla 5b   IVNSNTDMCTLVDGN*TAWEHQQCL*DWMSTFQLAYSLELSFGKITKQEVKD
AAV2            IVTSNTNMCAVIDGNSTTFEHQQPLQDRMFKFELTRRLDHDFGKVTKQEVKD
```


a)
Gray fox amdovirus
Mink amdovirus
Amdovirus
C.lanigera-2
O.degus-3
O.degus-1
Guinea pig EVE
C.lanigera-1
Wallaby EVE
O.degus-2
Opossum EVE
Southern hemisphere Parvovirus-related EVEs
0.1
Bufavirus-3
Bufavirus-2
Bufavirus-1
Bufavirus
Parvo/Bufavirus-related NS1
Porcine PV Tornau
Porcine PV
Feline PV
Mink PV
Canine PV
Rat PV 1a
PV H1
Rat MV 1c
Rat MV 1b
Rat MV 1a
PV LuIII
Hamster PV
Mouse PV 1b
Mouse PV 1a
Mouse MVb
Mouse MVa
Mouse MV I
Parvovirus
degenerated EVE
EVE with intact ORF
EVE with intact, transcribed ORF
b)
Snake parvovirus
Reptile
C.lanigera-4
Goose parvovirus
Duck parvovirus
Avian AAV-1 (VR-865)
Avian AAV-1
Avian
0.1
Guinea pig EVE
Wallaby EVE
Microbat EVE
Elephant EVE
Caprine AAV
Bovine AAV
O.degus-4
AAV4
AAV2
Baboon EVE
Mammalian
Dependovirus NS1

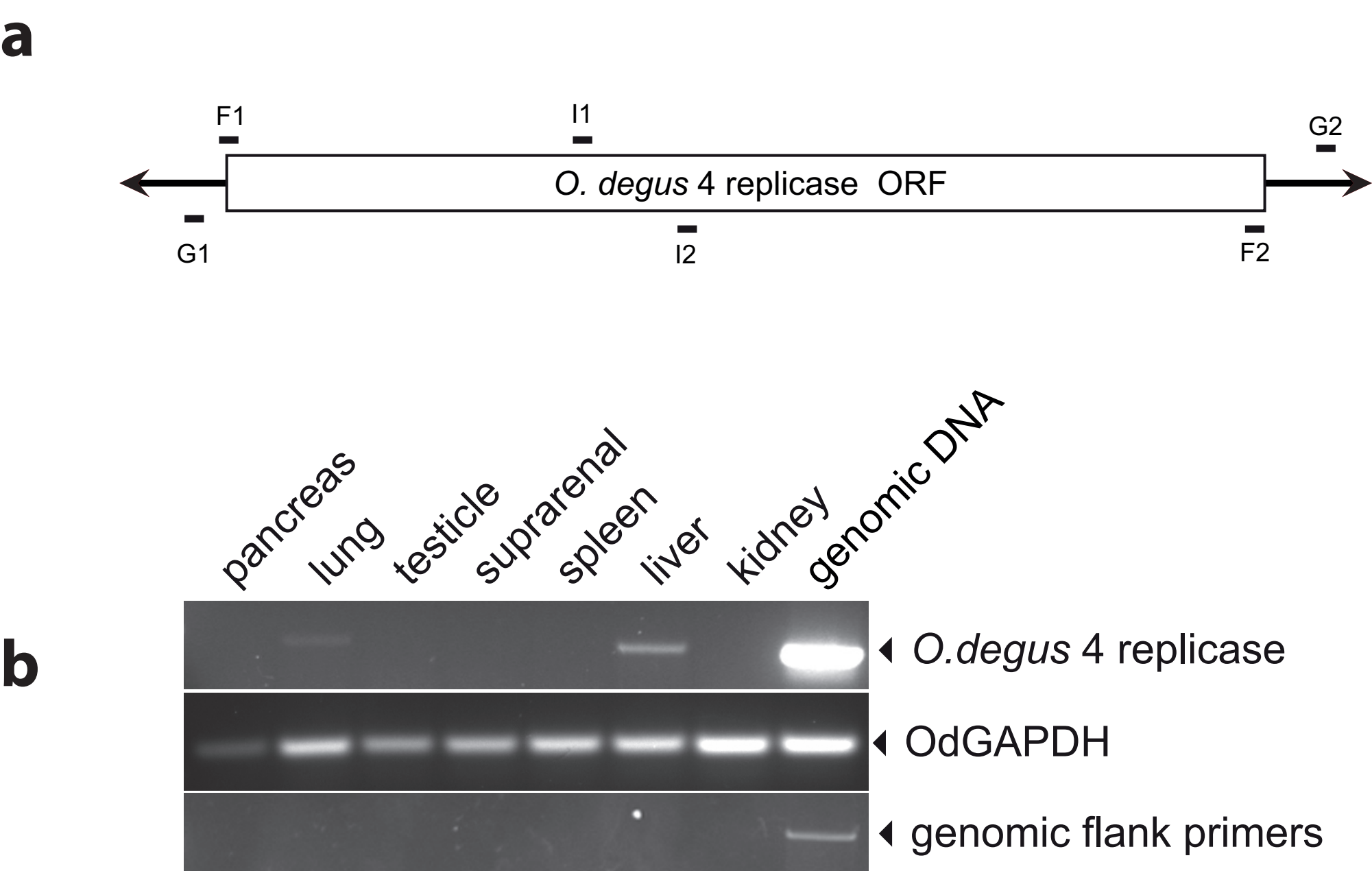

a
F1
I1
G2
O. degus 4 replicase ORF
G1
I2
F2
b
pancreas
lung
testicle
suprarenal
spleen
liver
kidney
genomic DNA
O.degus 4 replicase
OdGAPDH
genomic flank primers

Table 1. **Parvovirus EVEs in *C. lanigera and O.degus* genomes**

Element	Accession number	Orientation	Structure	Genus[1]	Match coordinates			
					Scaffold		*Viral genome*[2]	
					start	*end*	*start*	*end*
C.lanigera-1	JH721894.1	-ve	NS-VP	Parvo	14636996	14641202	1137	4398
C.lanigera-2	JH721993.1	-ve	NS-VP	Parvo	459340	462818	1065	4428
C.lanigera-3	JH721905.1	-ve	NS	Dependo	7599480	7600235	717	896
*C.lanigera-4**	JH721911.1	-ve	NS	Dependo	18094486	18095973	330	1684
C.lanigera-5a	JH721873.1	+ve	NS	Dependo	22669255	22669578	1428	1724
C.lanigera-5b	JH721873.1	+ve	NS	Dependo	22702412	22702712	1428	1724
C.lanigera-6	JH721896.1	-ve	VP	Dependo	13652894	13654833	4054	4407
O.degus-1	JH651603.1	+ve	NS-VP	Parvo	365906	369616	507	4443
O.degus-2	JH651624.1	-ve	NS-VP	Parvo	3857091	3858404	1089	2652
O.degus-3	JH651827.1	-ve	NS-VP	Parvo	1343057	1345620	295	3142
*O.degus-4**	JH651579.1	+ve	NS	Dependo	8427679	8429061	321	1784
O.degus-5	JH651577.1	+ve	VP	Parvo	12296699	12296983	4135	4419
O.degus-6	JH651549.1	-ve	VP	Dependo	16377593	16377937	4063	4401

Footnote: [1] Genus assignment based on phylogenetic analysis of the NS1 gene (see **Figure 2**). Dependo- = Dependovirus, Parvo- = parvovirus related EVEs [2] Viral genome coordinates based on pairwise alignment to genus-specific reference sequences; adeno-associated virus 2 (AF043303.1) for Dependoviruses, and mouse minute virus (J02275.1) for Parvovirus-like elements. Asterisks denote elements encoding intact open reading frames.